



# The hippocampus and inferential reasoning: building memories to navigate future decisions

Dagmar Zeithamova<sup>1,2,3†</sup>, Margaret L. Schlichting<sup>1,2†</sup> and Alison R. Preston<sup>1,2,3\*</sup>

<sup>1</sup> Center for Learning and Memory, The University of Texas at Austin, Austin, TX, USA

<sup>2</sup> Department of Psychology, The University of Texas at Austin, Austin, TX, USA

<sup>3</sup> Institute for Neuroscience, The University of Texas at Austin, Austin, TX, USA

## Edited by:

Joel Voss, Northwestern University, USA

## Reviewed by:

Joel Voss, Northwestern University, USA

Dharshan Kumaran, University College London, UK

## \*Correspondence:

Alison R. Preston, Center for Learning and Memory, The University of Texas at Austin, 1 University Station, C7000, Austin, TX 78712, USA.  
e-mail: [apreston@mail.clm.utexas.edu](mailto:apreston@mail.clm.utexas.edu)

## † Present address:

These authors equally contributed to this work.

A critical aspect of inferential reasoning is the ability to form relationships between items or events that were not experienced together. This review considers different perspectives on the role of the hippocampus in successful inferential reasoning during both memory encoding and retrieval. Intuitively, inference can be thought of as a logical process by which elements of individual existing memories are retrieved and recombined to answer novel questions. Such flexible retrieval is sub-served by the hippocampus and is thought to require specialized hippocampal encoding mechanisms that discretely code events such that event elements are individually accessible from memory. In addition to retrieval-based inference, recent research has also focused on hippocampal processes that support the combination of information acquired across multiple experiences during encoding. This mechanism suggests that by recalling past events during new experiences, connections can be created between newly formed and existing memories. Such hippocampally mediated memory integration would thus underlie the formation of networks of related memories that extend beyond direct experience to anticipate future judgments about the relationships between items and events. We also discuss integrative encoding in the context of emerging evidence linking the hippocampus to the formation of schemas as well as prospective theories of hippocampal function that suggest memories are actively constructed to anticipate future decisions and actions.

**Keywords:** hippocampus, inference, memory, encoding, retrieval, integration, flexibility

## INTRODUCTION

Many judgments and decisions in our everyday lives are not based on direct experience, but rather require inference based on knowledge acquired across multiple distinct experiences. For example, imagine you encounter an unfamiliar man leaving the house next door to walk his Great Dane. Because the house was recently sold, you might conclude that the man and his dog are your new neighbors. Several days later, you are in the park and see the same Great Dane again out for a walk, this time with a woman. From the knowledge acquired on these two separate occasions, you may infer a relationship between the man and woman; for instance, you may deduce that they are a couple and recently moved into the house next door with their Great Dane. Successful inferential reasoning may thus depend on our ability to recall detailed information from past events to determine how items experienced at different times are related. A growing body of literature indicates that such flexibility to combine experiences in novel ways to infer unobserved relationships between items or events crucially depends on the hippocampus.

The ability to infer a relationship between two previously experienced events is complex, involving several distinct operations. While some of these operations rely predominantly on brain structures outside the medial temporal lobe such as prefrontal cortex, others necessitate hippocampal processing. For successful

inference to ultimately take place, the arbitrary relations among previously unrelated elements within an event must be encoded (e.g., individual Great Dane–man and Great Dane–woman associations from the example above). To extract new information about the relationship between these events, encoded associations must be retrieved and then manipulated, recombined, and recoded based on their content to support the inference itself (man–woman). The precise contribution of hippocampus to each of these processes remains an area of active investigation and is a focus of this review. Notably, while these operations are all requisite for inference, the relative role of the hippocampus in these different processes—and the relative timing of each—may depend on the particular demands of the task at hand.

An important factor contributing to success in a variety of inference tasks is the nature of the underlying hippocampal memory representations. Decades of research have characterized how the hippocampus builds rich, detailed records of individual events, or episodic memories (Eichenbaum and Cohen, 2001; Squire et al., 2004; Eichenbaum et al., 2007; Preston and Wagner, 2007). Hippocampal memory representations are well-suited for the particular demands of inferential reasoning tasks, as the hippocampus is thought to discretely code multiple event elements in terms of their relationships to one another (Cohen and Eichenbaum, 1993; O'Reilly and Rudy, 2001). Such discrete

elemental coding allows hippocampal representations to be flexible, as relevant details of past events can be individually accessed as needed to support the types of novel decisions required in inferential reasoning.

Here, we discuss the developing literature linking hippocampal memory processes and representations to successful inference in a variety of tasks. By reviewing findings from both human and animal research, we highlight the hippocampal mechanisms that underlie mnemonic flexibility during both encoding and retrieval. We also argue that inferential reasoning provides a means of exploring the adaptive nature of memory, whereby memory representations are used to successfully negotiate current behavior and anticipate future decisions and actions.

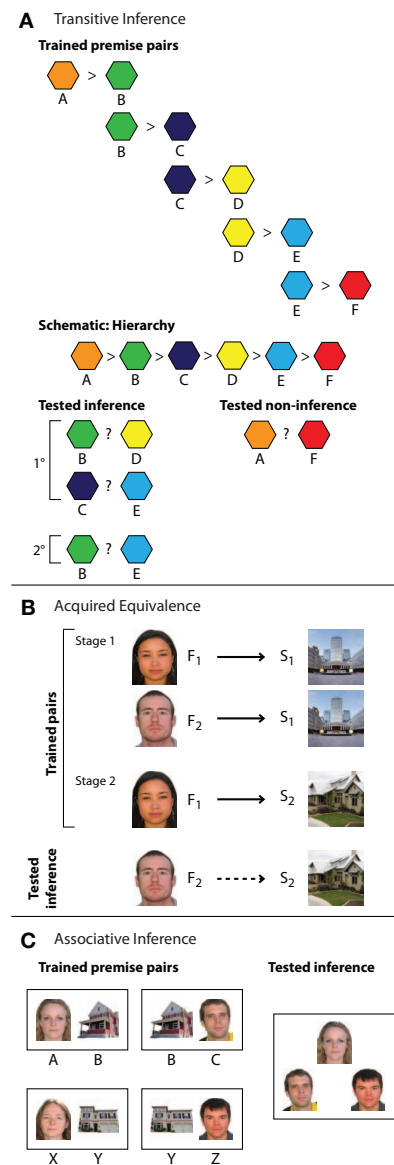
## INFERENTIAL REASONING TASKS

The novel expression of learned information has many forms, ranging from generalization of conditioned responses to novel stimuli in animals (Pavlov, 1927) to transfer of a learned task structure to new perceptual settings in humans (Kumaran et al., 2009). In this review, we focus specifically on the hippocampal mechanisms supporting the novel application of memory during inferential reasoning tasks that require judgments about the relationships between items experienced across discrete episodes.

Several paradigms have been used to study the role of the hippocampus in inferential reasoning (Figure 1), of which the most widely used in both animal (e.g., Dusek and Eichenbaum, 1997) and human research (e.g., Heckers et al., 2004) is the transitive inference task (Figure 1A). In this task, participants learn a set of overlapping premise relationships (e.g.,  $A > B$ ,  $B > C$ ,  $C > D$ ,  $D > E$ ,  $E > F$ ) via trial and error. During this initial training phase, participants learn to select the correct (i.e., reinforced) item. Training typically continues until a criterion level of performance on premise associations is reached.

Notably, multiple types of representations may support learning in this task. For instance, knowledge of reinforcement histories alone may guide memory for the end items of the hierarchy (A is always rewarded, F is never rewarded) and individual conditional associations may support memory for the inner pairs in the hierarchy (e.g., C is rewarded in the context of D, but not in the context of B). Alternatively, all items may be represented simultaneously as an ordered hierarchy of relationships ( $A > B > C > D > E > F$ ) that concisely represents trained associations as well as information about the relationships between items that were not directly trained (e.g.,  $B > D$ ).

To assess which types of representations support performance as well as how they may depend on the hippocampus, knowledge of the premise associations is tested together with novel, untrained combinations of items during the critical test phase. Novel test trials include inferential pairs with one degree of separation between items (e.g.,  $B ? D$ ,  $C ? E$ ) as well as pairs with two degrees of separation (e.g.,  $B ? F$ ), depending on the total number of items in the hierarchy. Novel non-inferential pairings consisting of the end items of the hierarchy (e.g.,  $A ? F$ ) are also tested. Critically, correct performance on the inferential test trials can only be achieved by considering the overarching hierarchy of relationships because both items (e.g., B and D)



**FIGURE 1 | Inference tasks. (A)** Transitive inference task with six elements. A set of overlapping training pairs forms an ordered hierarchy of relationships. Participants learn each individual training pair via feedback-based learning (e.g.,  $A > B$ ) and are then tested on novel inference and novel non-inference judgments. Items in inferential probe trials may be separated by one element in the hierarchy (e.g.,  $B ? D$ , indicated as 1°) or two elements (e.g.,  $B ? F$ , indicated as 2°). Novel non-inferential probes test knowledge of the relationship between the end items of the hierarchy ( $A ? F$ ). **(B)** Acquired equivalence task. In stage one of training, participants are trained via feedback to associate two faces ( $F_1$  and  $F_2$ ) with a particular scene ( $S_1$ ). In stage two, participants learn to select a second scene ( $S_2$ ) when cued with one of the faces ( $F_1$ ). Inference is then measured as the proportion of trials on which participants choose  $S_2$  when cued with  $F_2$ . The schematic depicts trained stimulus-response relationships (solid black arrows) and inferential relationships (dashed black arrows). **(C)** Associative inference task. Participants learn an overlapping set of associations (here, face-house associations), in which two stimuli (a man and a woman) are associated with a common third item (a house). Novel inference trials evaluate knowledge for the indirect relationship between items (who lives together in the same house).

are reinforced at the same rate during training.<sup>1</sup> In contrast, success on non-inferential test trials involving the end items can be based solely on the reinforcement histories (A is always reinforced, F is never reinforced) and does not require reference to the hierarchical relationships. This difference in task demands may be reflected by differing degrees of hippocampal recruitment for inferential and non-inferential test probes. Specifically, hippocampal engagement may be unique to inferential judgments that require reference to an ordered hierarchy of stimuli.

A second task used to examine the role of the hippocampus in inference is the acquired equivalence task (**Figure 1B**; Myers et al., 2003; Shohamy and Wagner, 2008). In this task, participants learn a set of stimulus–response associations via feedback training, such as learning to select Scene<sub>1</sub> when cued with Face<sub>1</sub>. In the first stage, the stimulus–response relationships are organized such that two distinct cue stimuli are associated with the same response (e.g., Face<sub>1</sub>–Scene<sub>1</sub>, Face<sub>2</sub>–Scene<sub>1</sub>). In the second training stage, additional information is learned about one of those cue items (e.g., Face<sub>1</sub>–Scene<sub>2</sub>) that can also be inferred to be true for the second, equivalent item (e.g., Face<sub>2</sub>). Thus, unlike the transitive inference task, the overlapping associations in the acquired equivalence task do not form a logical hierarchy. Inference in the acquired equivalence task is assessed by testing participants' knowledge of novel, untrained associations (Face<sub>2</sub>–Scene<sub>2</sub>) that can be inferred through transfer of learned associations (Face<sub>1</sub>–Scene<sub>2</sub>) to the equivalent item. While premise pairs may be encoded as inflexible stimulus–response associations, only representations that encode discrete event elements together with the relationships among them—such as those formed by the hippocampus—are thought to support the acquired equivalence judgment.

Another commonly used inferential reasoning paradigm is the associative inference task (**Figure 1C**; e.g., Bunsey and Eichenbaum, 1996; Preston et al., 2004). In associative inference, stimuli are organized into groups of three and presented to participants as overlapping pairs (e.g., AB, BC and XY, YZ) using either feedback or observational training. Inferential performance is then assessed by asking participants to make judgments about the relationship between elements of overlapping pairs that were not explicitly studied together (e.g., AC, XZ). As in other inferential tasks, the premise associations could be encoded as unitized representations, such as an “AB” unit during observational learning or an A–B stimulus–response association during feedback training. However, like acquired equivalence judgments, associative inference would be supported only by discrete elemental hippocampal representations that enable flexible access to individual event details.

Because the associative inference task can employ observational learning procedures, it provides an additional flexibility in research design. While training in the transitive inference and acquired equivalence tasks is typically limited to a small set of

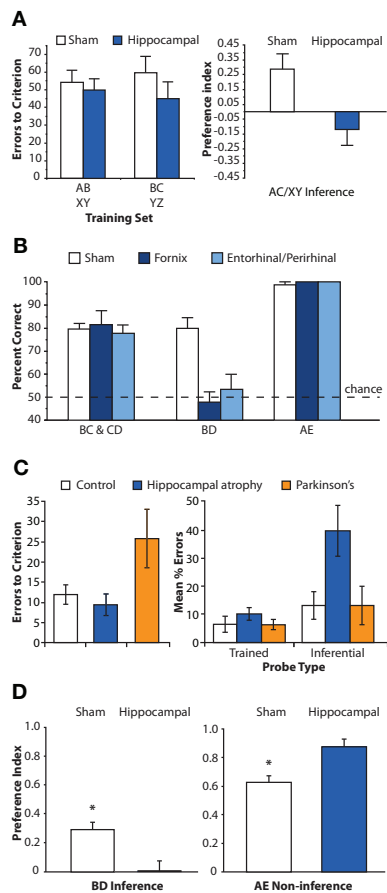
overlapping associations learned across multiple exposures, the associative inference task can be performed using a larger number of associations and single-trial learning procedures, wherein each trained association is only seen once during the learning phase (Zeithamova and Preston, 2010). Such rapid acquisition of arbitrary information is characteristic of the type of learning that occurs during daily episodic experiences and provides a means for studying how inference is performed with limited direct experience.

## CRITICAL ROLE FOR THE HIPPOCAMPUS IN INFERENTIAL REASONING

Converging evidence from animal and human research indicates that the hippocampus is necessary for successful performance in inferential reasoning tasks. In a series of animal lesion studies, Eichenbaum and colleagues trained rats on overlapping odor–odor associations using the associative inference (Bunsey and Eichenbaum, 1996) and transitive inference paradigms (Dusek and Eichenbaum, 1997). The hippocampal system was damaged prior to training, either by lesion to the hippocampus proper (Bunsey and Eichenbaum, 1996) or through disconnection of the hippocampus from its cortical and subcortical pathways (Dusek and Eichenbaum, 1997). In both tasks, hippocampal lesions impaired performance on inferential probe trials that tested knowledge of the untrained relationship between stimuli, while acquisition of the trained associations was unimpaired (**Figures 2A,B**). Similar impairments in transitive inference have been observed in non-human primates with lesions to the hippocampal system (Buckmaster et al., 2004). Notably, both lesioned rats (Dusek and Eichenbaum, 1997) and lesioned monkeys (Buckmaster et al., 2004) performed perfectly on novel non-inferential test trials that probed knowledge about the relationship between end items of the hierarchy (e.g., A ? F), suggesting that making judgments about novel combinations of familiar items does not itself require hippocampal processing. Rather, these findings indicate that the hippocampus plays an essential role in judgments that require the flexible manipulation of learned relationships among items when simple comparisons of reinforcement history do not suffice.

Evidence for an essential role of the hippocampus in the acquired equivalence task is somewhat inconsistent across species. Rodent research has shown impaired performance on a spatial variant of the acquired equivalence task in animals with lesions to entorhinal cortex but not in animals with hippocampal lesions (Coutureau et al., 2002). In contrast, neuropsychological research in humans indicates that hippocampal lesions critically impact performance in the acquired equivalence task (Myers et al., 2003). Patients with hippocampal atrophy acquire stimulus–response associations at a rate similar to control patients and demonstrate intact memory for these trained associations during the test phase (**Figure 2C**). However, these patients are impaired on inferential judgments testing knowledge of the untrained, equivalent relationships between items. In contrast, due to basal ganglia damage, Parkinson's disease patients demonstrate slower acquisition of the premise associations during the feedback-based learning phase, but intact performance on both trained and inferential probes at test. This double dissociation demonstrates the critical,

<sup>1</sup> An alternative account of the transitive inference task proposes that inferential probe trials can be solved based on the individual reinforcement histories of trained stimuli and does not necessitate the formation of a hierarchical representation (Frank et al., 2003).



**FIGURE 2 | Lesion and neuropsychological studies assessing the critical role of hippocampus in inference. (A)** Associative inference task. Left panel: mean number of errors to criterion on training of two sets of overlapping associations (e.g., AB, BC). Right panel: Inference performance as measured by a preference index for indirectly related item (e.g., selecting C when cued with A). White bars denote sham operated control rats; blue bars denote hippocampally lesioned rats. Hippocampally lesioned rats learn individual relationships between item pairs at a rate similar to control rats, but fail on the inference test. Adapted by permission from Bunsey and Eichenbaum (1996), copyright 1996 Macmillan Publishers Ltd. **(B)** Transitive inference task. Rats with lesions disconnecting the hippocampus from its subcortical (fornix, dark blue bars) or cortical (entorhinal/perirhinal, light blue bars) target structures performed similarly to sham operated control rats (white bars) on trained associations (BC, CD) and novel non-inferential probe trials (AE). However, lesioned rats were severely impaired on inferential probe trials (BD). Adapted by permission from Dusek and Eichenbaum (1997), copyright 1997 National Academy of Sciences, USA. **(C)** Mean performance of control participants (white bars), patients with hippocampal atrophy (blue bars) or Parkinson's disease patients (orange bars) in an acquired equivalence task. Patients with hippocampal atrophy reached a criterion level of performance at a rate similar to control participants. In contrast, Parkinson's disease patients required more extensive training. However, at test, hippocampal patients were severely impaired on inferential probe trials relative to both control participants and patients with Parkinson's disease. Adapted by permission from Myers et al. (2003), copyright 2003 MIT Press. **(D)** Post-training hippocampal lesions (blue bars) impaired transitive inference judgments (BD) in mice, but enhanced performance on novel non-inferential probe trials (AE) involving the end items of the hierarchy relative to sham operated animals (white bars). Adapted by permission from DeVito et al. (2010a), copyright 2009 Wiley-Liss, INC.

specialized role of the hippocampus in decisions that require the flexible application of learned knowledge to novel situations in humans.

These initial neuropsychological and animal lesion studies provide substantial evidence for the critical role of the hippocampus in inferential reasoning across a variety of experimental paradigms. One finding common across these experiments is intact learning of the explicitly trained associations despite hippocampal damage. This may suggest that while the hippocampus is not required for the acquisition of individual premise associations during encoding (when trained across multiple repetitions using feedback-based learning procedures), it is essential for retrieving and recombining event elements during inference tests. An alternative possibility is that hippocampal lesions induce changes in encoding strategy, resulting in a different representational form for the trained associations that does not allow for the flexible recombination of information during inference itself (Eichenbaum et al., 1996; Frank et al., 2003).

From these initial lesion and neuropsychological studies alone, it is not possible to determine the precise stage—encoding or retrieval—or the precise mechanism of hippocampal involvement in inference, as hippocampal damage was present prior to the initial training phase. More recent findings indicate that multiple hippocampal mechanisms contribute to successful inferential reasoning, including processes engaged during initial encoding, flexible retrieval, and post-encoding sleep. We discuss these ideas by reviewing evidence from functional magnetic resonance imaging (fMRI) studies in humans demonstrating changes in hippocampal activation during different phases of inferential reasoning tasks along with convergent findings from computational modeling and more recent animal lesion studies.

## HIPPOCAMPAL RETRIEVAL PROCESSES THAT SUPPORT INFERRENTIAL REASONING

Inferential reasoning is traditionally thought of as a logical process in which novel relationships are deduced from knowledge about premise relationships. In the example with the man, woman, and Great Dane, you have no direct knowledge about either the relationship between the man and the woman or where the woman lives. When faced with a misplaced piece of mail addressed to the house next door, you might retrieve information acquired during two previous events—that “the man in the house next door owns the Great Dane” and “the Great Dane belongs to the woman”—and recombine that knowledge to conclude that “the woman is my new neighbor.” In doing so, you determine that you can deliver the mail to her the next time you see her in the neighborhood. In this way, inferential reasoning is accomplished at the time of retrieval when faced with a novel judgment.

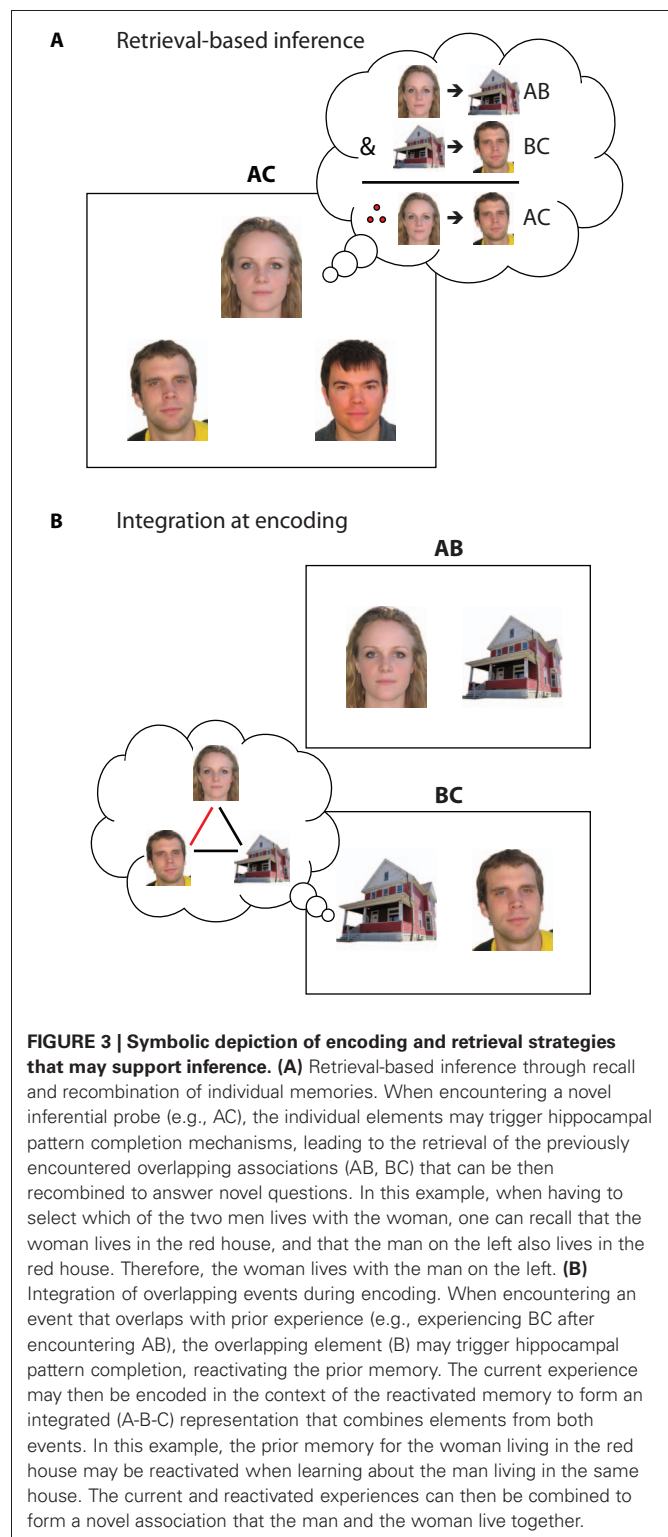
While inferential reasoning itself has not been traditionally conceptualized as a function of hippocampus, recent research highlights the hippocampal role in several of the processes contributing to this ability. In the example above, successful inference requires the initial encoding of associations, retrieval of these associations through individual elements when faced with a misplaced envelope, and subsequent recombination of information to yield a solution. Such retrieval-based processes that allow for the flexible use of previous experience are hypothesized to rely



on hippocampal memory representations that code event details individually in terms of their relationships to one another (Cohen and Eichenbaum, 1993; O'Reilly and Rudy, 2001). This discrete, relational coding allows event elements to be individually addressable and retrievable from partial input—a process often referred to as *pattern completion* (Marr, 1971; McClelland et al., 1995). In inference tasks used in the laboratory, novel test probes (e.g., AC test trials in associative inference) would trigger hippocampal pattern completion, leading to the retrieval of previously encountered, overlapping memories (i.e., A was paired with B, B was paired with C) that could then be recombined to support successful inferential judgments (**Figure 3A**).

Several human neuroimaging studies have provided evidence that the hippocampus plays an important role in successful inference at the time of retrieval (Heckers et al., 2004; Preston et al., 2004; Zalesak and Heckers, 2009). In one such study, participants were trained on overlapping face–house associations (AB, BC) and non-overlapping face–face associations (DE) in a modified version of the associative inference paradigm (Preston et al., 2004). Participants were then tested on the learned associations (AB, BC, DE) as well as novel face–face associations (AC) that required mediation through explicitly learned common houses (Figure 4A). While a region in posterior hippocampus was engaged for all associative retrieval trials, bilateral anterior hippocampus was uniquely engaged during inferential (AC) memory probes. This pattern of response was unique to the anterior hippocampus, providing support for its role in the flexible use of memory at retrieval.

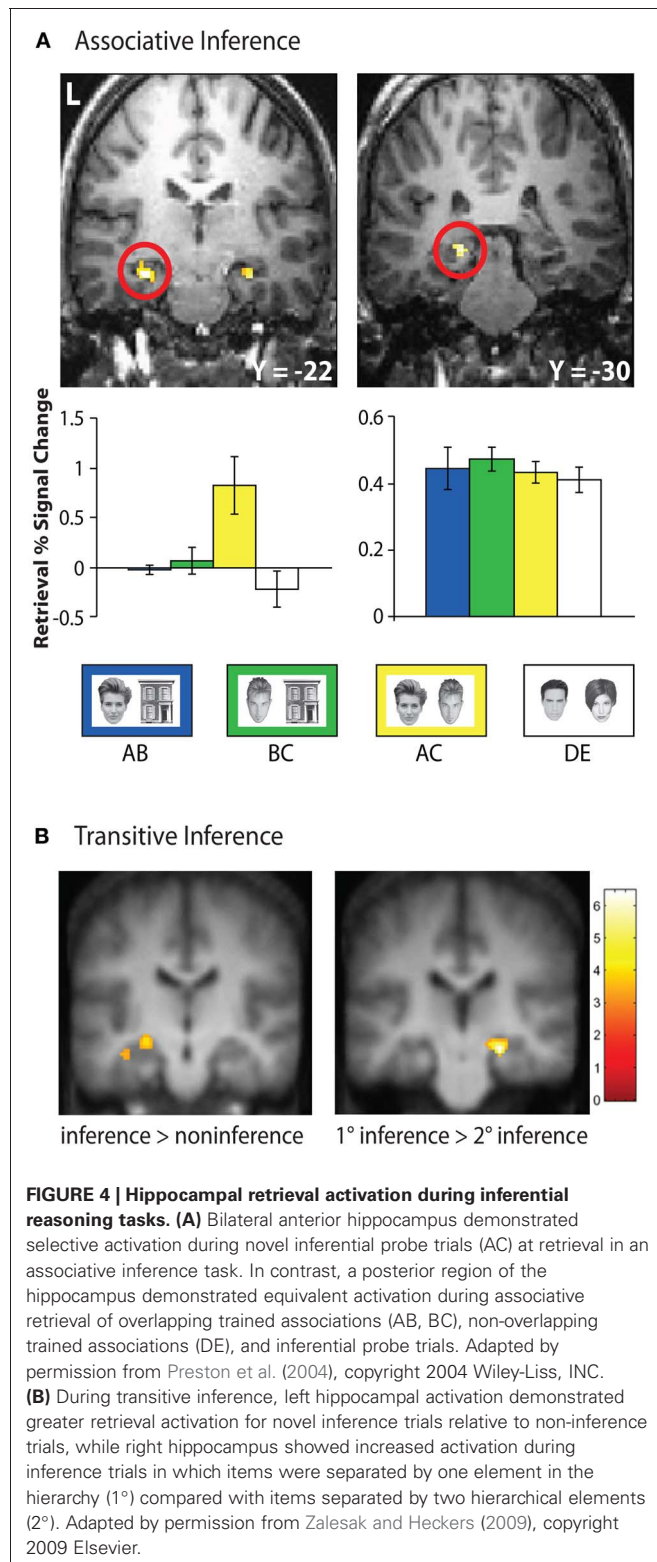
Hippocampal engagement during retrieval-based inference has also been observed in transitive inference tasks (Heckers et al., 2004; Zalesak and Heckers, 2009). In one such study, participants were presented with pairs of stimuli and trained to select the rewarded item from each pair (e.g.,  $A > B$ ). Two sets of pairs were used: overlapping pairs ( $A > B$ ,  $B > C$ ,  $C > D$ ,  $D > E$ ) forming an ordered hierarchy and non-overlapping pairs ( $a > b$ ,  $c > d$ ,  $e > f$ ,  $g > h$ ) with no hierarchical structure (Heckers et al., 2004). Novel test probes consisting of the elements from overlapping pairs (e.g.,  $A ? C$ ,  $B ? D$ ) produced greater hippocampal engagement than did novel test probes comprised of items from non-overlapping associations (e.g.,  $c ? f$ ). However, this study did not directly address the role of hippocampus in inferential test trials, as there was no comparison made between inferential probe trials requiring reference to the stimulus hierarchy (e.g.,  $B ? D$ ) and non-inferential probe trials containing the hierarchy's end items (i.e.,  $A ? E$ ). In a follow-up experiment (Zalesak and Heckers, 2009), participants were trained on a six-element hierarchy and tested on inferential relationships between inner members of the hierarchy (i.e.,  $B ? D$ ,  $C ? E$ ,  $B ? E$ ) and non-inferential pairs including the end items (e.g.,  $A ? D$ ,  $C ? F$ ) that could be solved based solely on knowledge of the reinforcement histories. Left hippocampus showed greater activation during retrieval of inferential relative to non-inferential probe trials (**Figure 4B**). Additionally, the degree of hippocampal activation during inferential probe trials was related to the relative distance between items in the transitive hierarchy. Greater activation was observed in right hippocampus for inferential judgments with one degree of separation between probe items (i.e.,  $B ? D$ ,  $C ? E$ ) relative to inferential judgments



**FIGURE 3 | Symbolic depiction of encoding and retrieval strategies that may support inference. (A)** Retrieval-based inference through recall and recombination of individual memories. When encountering a novel inferential probe (e.g., AC), the individual elements may trigger hippocampal pattern completion mechanisms, leading to the retrieval of the previously encountered overlapping associations (AB, BC) that can be then recombined to answer novel questions. In this example, when having to select which of the two men lives with the woman, one can recall that the woman lives in the red house, and that the man on the left also lives in the red house. Therefore, the woman lives with the man on the left. **(B)** Integration of overlapping events during encoding. When encountering an event that overlaps with prior experience (e.g., experiencing BC after encountering AB), the overlapping element (B) may trigger hippocampal pattern completion, reactivating the prior memory. The current experience may then be encoded in the context of the reactivated memory to form an integrated (A-B-C) representation that combines elements from both events. In this example, the prior memory for the woman living in the red house may be reactivated when learning about the man living in the same house. The current and reactivated experiences can then be combined to form a novel association that the man and the woman live together.

with two degrees of separation between probe items (i.e.,  $B ? E$ ), suggesting that hippocampal activation tracks the degree of relational processing required for successful inference.

While these human fMRI studies demonstrate that the hippocampus is engaged during inferential judgments, they cannot



demonstrate whether such engagement is *necessary* for successful performance. In a recent animal lesion study, selective hippocampal damage produced after acquisition of overlapping memories severely impaired performance on inference judgments in a

transitive inference task, providing direct evidence for the essential role of the hippocampus beyond the initial training phase (**Figure 2D**; DeVito et al., 2010a). Interestingly, animals with hippocampal lesions performed more accurately than did control animals on the non-inferential probe trials involving end items. These data show that animals with hippocampal lesions relied primarily upon the reinforcement histories of individual stimulus elements (most salient for end items) rather than the relative contingencies of reinforcement between items. Presumably, post-encoding hippocampal lesions specifically eliminated memory for the relationships between items that would be needed for successful inferential performance, while leaving intact information about the reinforcement patterns of individual stimuli.

Collectively, these findings indicate that hippocampal memory representations are accessed during inferential judgments. However, additional evidence suggests that other brain regions are recruited in concert with the hippocampus in service of successful inference. Activation in the dorsolateral and medial prefrontal cortex has been observed during transitive inference in humans (Acuna et al., 2002), and lesions to medial prefrontal cortex produce selective impairments on the critical inference judgments in the transitive inference (DeVito et al., 2010b) and acquired equivalence tasks (Iordanova et al., 2007). Furthermore, activation in prefrontal cortex and medial temporal lobe (MTL) structures is more tightly coupled during inferential retrieval judgments relative to memory judgments for directly learned information in an associative inference task, with activation in inferior frontal gyrus and medial prefrontal cortex further differentiating between correct and incorrect inference performance (Zeithamova and Preston, 2010).

The respective roles of hippocampus and prefrontal cortex in inferential retrieval have yet to be determined. One possibility is that the hippocampus supports memory for directly experienced associations, while structures in prefrontal cortex sub-serve relational reasoning processes (Robin and Holyoak, 1995; Christoff et al., 2001) whereby the contents of hippocampal representations are manipulated and recombined to support the inference itself. Some evidence for this view comes from a non-mnemonic inferential reasoning task in which all premise relationships were concurrently displayed along with the probe judgment (Wendelken and Bunge, 2010). In this study, the hippocampus was similarly engaged during both probe trials requiring consideration of a single premise relationship and probe trials requiring inference across multiple premise relationships. However, prefrontal cortex was preferentially engaged during inferential judgments. In this non-mnemonic inference task, the hippocampus may be processing individual relationships among premise elements, an operation required for both probe trial types. Only those probe trials requiring inference across multiple premises necessitate manipulation and recombination through prefrontal processing to support inference. An alternative possibility is that hippocampus may also contribute to the actual manipulation of relational information (Hannula et al., 2006; Hannula and Ranganath, 2008), creating novel combinations of elements extending beyond prior experience. Findings demonstrating hippocampal engagement unique to inferential judgments (Preston et al., 2004; **Figure 4A**) may reflect such processes. While the respective contributions of

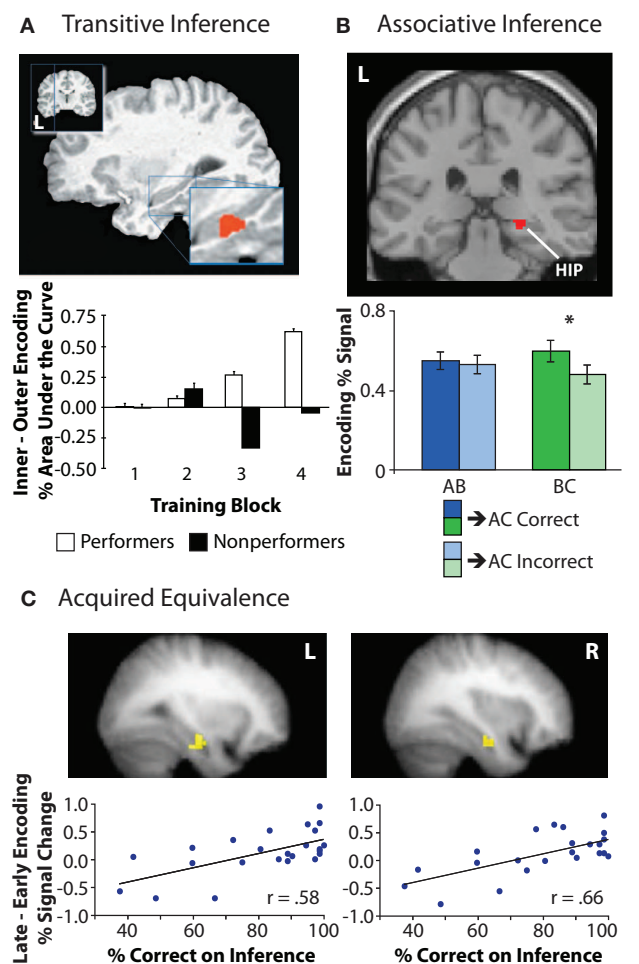
the hippocampus and prefrontal regions in inferential reasoning remains an active area of study, there is substantial evidence suggesting that these regions form an integrated network for relational processing (DeVito et al., 2010b; Ranganath, 2010) that is engaged during inferential reasoning.

### HIPPOCAMPAL ENCODING PROCESSES THAT SUPPORT INFERRENTIAL REASONING

Initial studies of inference primarily focused on the role of hippocampus in flexible retrieval processes in which novel probes (e.g., B ? D in transitive inference) trigger recall of directly experienced memories (B > C, C > D), with inference (B > D) being supported by flexible recombination of recalled memories. This focus on retrieval-based processes describes how memories are recombined or modified *after* they are initially encoded. However, more recent neuroimaging studies support the notion that hippocampal encoding plays an equally important role in successful inference.

In one of the first human neuroimaging studies on inference, Nagode and Pardo (2002) used positron emission tomography (PET) to examine the relationship between hippocampal encoding activation and successful transitive inference. Greater hippocampal activation was observed during the intermediate relative to initial stages of training, suggesting experience-dependent changes in hippocampal recruitment during encoding. In a second experiment, participants were first trained on non-overlapping portions of the transitive hierarchy (A > B, C > D, E > F, G > H) followed by training on the “bridging,” or overlapping, pairs (B > C, D > E, F > G) that connected the initially learned associations. Hippocampal activation was greater during training of the bridging pairs relative to training on the non-overlapping pairs in the hierarchy. A similar pattern of findings was shown by Greene and colleagues, who scanned participants during both training and test phases of the transitive inference task using fMRI (Greene et al., 2006). During the latter part of the learning phase, left hippocampal activation was greater for inner training pairs (B > C, C > D) relative to outer training pairs (A > B, D > E), with this difference in activation correlating with individual differences in inferential performance (Figure 5A). Greater test-phase activation was also observed in hippocampus for inferential test probes (B ? D) among those participants who were successful at inference (“performers”) relative to unsuccessful participants (“non-performers”), suggesting that the way in which associations are initially encoded relates to the ability to later use them flexibly. Together, the results of these neuroimaging studies demonstrate a contribution of hippocampal encoding processes to novel inferential judgments.

Animal research is also consistent with the idea that hippocampal encoding processes are important for successful inference. As previously discussed, hippocampal damage that occurs after initial learning eliminates memory representations that code the relationships among elements in a transitive hierarchy, while leaving intact information about the reinforcement histories of individual items (DeVito et al., 2010a). Presumably, animals with post-training lesions demonstrate significant impairments on inference tests due to loss of the representations of critical relationships among event elements formed by the hippocampus



**FIGURE 5 | Hippocampal encoding activation during inferential reasoning tasks. (A)** Left hippocampal activation increased across training block for inner pairs in the transitive hierarchy (B > C) relative to outer pairs (A > B), but only for those participants who were successful on the inferential test. Adapted by permission from Greene et al. (2006), copyright 2006 MIT Press. **(B)** In an associative inference task, right hippocampal activation during encoding of overlapping associations (BC) was greater for trials in which the corresponding inference judgment (AC) was later correct relative to trials on which the inference judgment was later incorrect. Hippocampal activation during initially acquired associations (AB) was not related to subsequent inferential performance. Adapted from Zeithamova and Preston (2010). **(C)** Activation in left and right hippocampus during the training phase of an acquired equivalence task was correlated with individual differences in inference performance. Specifically, increases in bilateral hippocampal activation from the early to late portion of the training phase were associated with superior performance on inferential probe trials. Adapted by permission from Shohamy and Wagner (2008), copyright 2008 Elsevier.

during encoding. Furthermore, recent animal research indicates that hippocampal lesions performed prior to learning produce even greater deficits in inferential performance than do post-training lesions (van der Jeugd et al., 2009), highlighting the critical role of hippocampus during the encoding process.

While the above discussed human neuroimaging and animal lesion studies established the importance of the hippocampal encoding processes in successful inference, the



precise mechanisms by which hippocampus contributes to performance cannot be determined from these data alone. Two hippocampal encoding mechanisms have been proposed to underlie subsequent inference: (1) elemental encoding of individual premise associations and (2) integrative encoding. As previously expressed, elemental encoding is critical to successful inference and refers to the initial encoding of experience such that memories can be later accessed through individual event details. Such elemental representations formed by hippocampus during encoding are essential for retrieval-based inference processes in that they enable access of necessary details when faced with a novel judgment about items not directly experienced together.

In addition to the flexible, elemental encoding of individual associations, compelling new evidence suggests that the hippocampus may also support inferential judgments by dynamically integrating newly encountered information into existing memory networks at the time of learning—a process termed integrative encoding (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010). According to this hypothesis, the hippocampus not only forms relationships among elements within an individual experience, but also links elements across discrete experiences.

During integrative encoding, new experiences are not only encoded in the context of presently available information, but also in the context of internally generated memory representations of prior overlapping events (**Figure 3B**). Through reactivation of previous experience (Eichenbaum, 2000; O'Reilly and Rudy, 2001; Shohamy and Wagner, 2008), hippocampal processing may allow for integration of information across distinct experiences in anticipation of future use. This constructive, or prospective, nature of memory (Klein et al., 2002; Buckner, 2010; Addis and Schacter, 2011) dates back to Tolman's concept of a "cognitive map" (Tolman, 1948) and has been proposed as a key mechanism underlying successful inferential reasoning (O'Reilly and Rudy, 2001; Shohamy and Wagner, 2008; Zeithamova and Preston, 2010). As related memories have already been integrated during encoding, this mechanism—in contrast to flexible recombination at retrieval—predicts that inference judgments may not require any additional processing during test.

For example, let us return to the scenario with your new neighbors, the man, woman, and Great Dane. When you initially see the man leaving the house next door with his Great Dane, you form a memory for the event that represents the relationship between the man, the dog, and the house. Upon seeing the Great Dane a second time with the woman, the familiar element (the Great Dane) may serve as a cue for hippocampal pattern completion, leading to the reactivation of your prior experience with the dog. The new event (the woman walking the Great Dane) is then encoded in the presence of the reactivated information about your first experience with the dog. In this way, a link between the man, the woman, and the house next door can be formed during encoding, despite the fact that you have never seen the woman with the man or at the house next door. Therefore, when you receive the misplaced piece of mail addressed to the house next door, no new recombination of information is required; rather, you can directly retrieve your memory that the man and the woman are your new neighbors and determine that you can deliver the mail to either of them when you see them around the neighborhood.

Importantly, elemental encoding of individual associations and integrative encoding are not mutually exclusive. Rather, integration of new information into an existing memory depends on elemental encoding of the initial memory such that it can be reactivated when the overlapping element (the Great Dane) is encountered again in the second episode.

Recent animal and human research has demonstrated reactivation of prior events during new learning. For instance, electrophysiological studies in rodents have shown hippocampally mediated replay of prior event sequences in new spatial contexts (Karlsson and Frank, 2009) and sequential activation of hippocampal place cells for never-experienced spatial trajectories that represent a shortcut through a well-learned environment (Gupta et al., 2010). Furthermore, in environments with overlapping elements, individual hippocampal neurons demonstrate experience-dependent generalized firing patterns that respond in multiple similar locations (Singer et al., 2010) or to the overlapping features themselves (Wood et al., 1999). Such generalized firing patterns suggest that hippocampal neurons develop representations that code the similarities between events. By representing features common to multiple events similarly, hippocampal codes can capture regularities shared across different experiences and, in doing so, may act as "nodes" that link distinct behavioral episodes (Eichenbaum et al., 1999). Converging findings from human neuroimaging research have demonstrated hippocampally mediated reactivation of prior experience during encoding of new overlapping events that was associated with greater retention of originally learned information (Kuhl et al., 2010). These findings demonstrate that reactivating memories during new learning helps reduce forgetting of past events, and it is possible that a similar mechanism underlies successful inferential reasoning.

The notion that related events are integrated during encoding is consistent with the symbolic distance effect, which refers to increased accuracy and decreased reaction time for judgments comparing items farther apart in a stimulus hierarchy. The symbolic distance effect is often (Frank et al., 2005; Zalesak and Heckers, 2009), although not universally (Moses et al., 2006), observed in transitive inference tasks. Integration during encoding would promote the formation of a single hierarchical representation of overlapping stimuli ( $A > B > C > D > E > F$ ) rather than the formation of representations of each individual pair ( $A > B$ ,  $B > C$ ,  $C > D$ ,  $D > E$ ,  $E > F$ ). If the transitive inference task was solved by retrieval and recombination of individual learned relationships at test, it should be easier to infer the relationship between items with one degree of separation (e.g.,  $B > D$ ) than between items with two degrees of separation (e.g.,  $B > E$ ), as the former requires retrieval of just two representations ( $B > C$  and  $C > D$ ) while the latter requires retrieval of three ( $B > C$ ,  $C > D$ , and  $D > E$ ). A single hierarchical representation has the advantage of directly encoding the relationships between distant stimuli along with the directly learned relationships, making indirectly learned relationships readily available at the time of test. In contrast to the retrieval-based mechanism, this representational structure would predict that inferences about items far apart in the hierarchy (e.g.,  $B > E$ ) would be easier than inferences for items that are close together (e.g.,  $B > D$ ).



because their relative positions in the hierarchy are more distinct.<sup>2</sup>

Recent neuroimaging studies provide more direct evidence for an integrative encoding process in hippocampus in acquired equivalence (Shohamy and Wagner, 2008) and associative inference paradigms (Zeithamova and Preston, 2010). In the acquired equivalence task, increases in hippocampal activation across the learning phase were associated with individual differences in inferential performance, even when accounting for performance differences on trained associations (**Figure 5C**). Moreover, no difference in test-phase activation between inferential and trained probes was observed, and reaction times for inferential probe trials and trained associations at test did not differ for successful participants (Shohamy and Wagner, 2008). These findings suggest that successful performance in the acquired equivalence task is dependent upon the formation of the inferential relationships during encoding that are immediately available when probed with novel combinations of items at test. Hippocampal encoding activation also predicted trial-by-trial success in the associative inference task (Zeithamova and Preston, 2010). Specifically, trial-by-trial measures of hippocampal activation during encoding of overlapping associations (BC), but not initially acquired associations (AB), differentiated between subsequently correct and incorrect inferential judgments (AC; **Figure 5B**). Importantly, while both hippocampal and prefrontal responses are implicated in successful inference performance at the time of test, this pattern of encoding activation is only observed in the hippocampus and surrounding MTL cortex. Thus, converging evidence from these different paradigms suggests that successful inference can be supported by a hippocampus-specific encoding mechanism whereby overlapping experiences are integrated into a network of related memories as they are learned.

Current computational models of hippocampal function emphasize a specific role for the CA<sub>3</sub> region in the formation of integrated relational memory networks (Wallenstein et al., 1998). According to such models, CA<sub>3</sub> neurons develop “context fields” during learning that bind together elements within a single event. Through such binding, CA<sub>3</sub> context fields respond preferentially to temporally contiguous stimuli or events. This mechanism that binds together items in the same sequential context may also provide a potential neural substrate for the integration of experiences that share common features beyond temporal context. In simulation experiments, the formation of context fields in CA<sub>3</sub> models during the acquisition phase of the associative inference task led to correct performance on the inferential probe trials during test (Wallenstein et al., 1998). In contrast, CA<sub>3</sub> models that did not develop context fields failed on associative inference trials, despite successful acquisition of the trained associations. This finding suggests that CA<sub>3</sub> binding processes are critical to the formation of integrated memories and successful inference.

Future animal research or human neuroimaging studies utilizing high-resolution fMRI techniques (Zeineh et al., 2003; Bakker et al., 2008; Preston et al., 2010) may provide additional insight into the precise hippocampal mechanisms that support the type of mnemonic flexibility required in inferential reasoning. For an extended discussion of computational perspectives on the role of hippocampus in inference, see the review by Kumaran in this special topic (Kumaran, 2012).

It is noteworthy that a direct contradiction exists among neuroimaging studies of inferential reasoning: while several studies have demonstrated enhanced hippocampal engagement during inferential reasoning probes (Heckers et al., 2004; Preston et al., 2004; Zalesak and Heckers, 2009; Wendelken and Bunge, 2010; Zeithamova and Preston, 2010), others have failed to find evidence for retrieval-based hippocampal processing during inference, instead linking inferential ability primarily to encoding processes (Nagode and Pardo, 2002; Shohamy and Wagner, 2008). While it is important to not over-interpret the null test-phase findings in these studies, several possible explanations may be worth exploring in future research that may help make sense of these apparent empirical contradictions. For example, the number of training trials during the learning phase of inference tasks may have a major influence on the type of hippocampal representation recruited in service of successful inference. In many cases, a limited set of overlapping associations (e.g., five in transitive inference) are trained across many repetitions (e.g., Nagode and Pardo, 2002). In other cases, a large number of overlapping associations are learned in a single exposure (Zeithamova and Preston, 2010) or relatively few repetitions (Preston et al., 2004). Integrative encoding may fully support inference during learning of a limited set of experiences across many repetitions, as there are multiple opportunities to build and strengthen the links between different trained associations as they are learned. In contrast, such hippocampally mediated integration may not be sufficient in tasks that utilize large stimulus sets and single-trial learning procedures, leading to the additional recruitment of retrieval-based hippocampal mechanisms. For example, both an integrative encoding signature and increased retrieval activation for inferential probes were observed in a single trial associative inference task (Zeithamova and Preston, 2010), suggesting that both processes may be recruited in a single task. Other task demands may similarly influence when one hippocampal mechanism—integration during encoding or recombination at retrieval—is favored over another, including feedback-based vs. observation learning and interleaved vs. blocked presentation of training associations. Understanding how the dynamics of the task influence the recruitment of the hippocampus at different stage of learning may provide key insights into computational properties of the hippocampus and its functional role in inferential reasoning.

## INFERENCE AND SCHEMAS

The notion that the hippocampus contributes to inferential reasoning by integrating new experiences into existing memory networks to form links between distinct events is conceptually related to an emerging body of literature on the role of the hippocampus in the formation of schemas. Schemas are knowledge

<sup>2</sup>The symbolic distance effect is also consistent with the value transfer account of the transitive inference task, which proposes knowledge of individual reinforcement histories underlies performance. Notably, this alternative account also argues against a retrieval-based mechanism for inference (von Fersen et al., 1991; Frank et al., 2003).

frameworks that capture regular patterns in the environment by abstracting information across experiences (Bartlett, 1932) and represent features common to multiple different events while discarding idiosyncratic details. For example, a “restaurant schema” may contain commonly experienced elements such as sitting at a table, ordering from a menu, and paying the bill, but not one-time elements such as the waiter spilling water on you.

Schemas guide behavior by providing a set of expectations for a given experience. Like integrated memory representations, schemas also contain information derived from multiple events that may support inferential decisions. Specifically, schemas represent relationships between elements commonly associated with certain types of situations, despite the fact that these elements have not necessarily been experienced together. Moreover, encoding new events in the context of a reactivated schema may provide an additional mechanism for inferential reasoning. For example, a person may come to your table at the end of your meal and inquire about the quality of the food and service. In the absence of an introduction, you may infer that this person is the owner or manager of the restaurant because your restaurant schema contains information about who is likely to ask for feedback about your dining experience. Like the integrative encoding processes discussed in the previous section, schemas build knowledge representations from multiple individual events and may thus involve neural mechanisms similar to the hypothesized integrative encoding processes that support inference.

Few studies to date have directly explored the role of the hippocampus in the formation of schemas. One recent neuroimaging study in humans (Kumaran et al., 2009) revealed hippocampal-ventromedial prefrontal cortex (VMPFC) interactions during application of a task schema; only hippocampus was associated with speeded transfer to a perceptually novel task with a similar structure. Schema-dependent speeded learning has also been observed in rodent research, which demonstrated that reactivation of an existing task schema (in this case, a familiar spatial layout) allowed for rapid acquisition of new flavor–place associations in a single trial (Tse et al., 2007, 2011). Without an existing schema, such associative learning required repeated training across multiple weeks. Importantly, rats with hippocampal lesions fail to show facilitated learning of new information in the presence of reactivated schemas, highlighting a critical role for this region in the rapid incorporation of new information into existing knowledge frameworks.

While direct evidence linking schema formation and updating to integrative encoding processes observed in inference tasks is lacking, the striking similarities between findings from these two literatures provide strong evidence that the hippocampus plays a unique role in binding processes that integrate information across distinct events. Furthermore, several inference studies have examined well-learned knowledge structures by extensively training participants on a set of associations before introducing overlapping events. This potentially schema-like knowledge could then support the rapid encoding of new overlapping information, as observed in spatial learning paradigms in rodents. However, whether hippocampal engagement in inference tasks reflects a similar or precisely the same representational mechanism as that employed during schema formation is yet to be seen.

While one important characteristic of schemas is the loss of idiosyncratic details that code the differences among events, it remains unknown whether the same is true of integrated memory representations. Anecdotal evidence from the acquired equivalence paradigm suggests that some event details may also be lost during integration, as participants fail to recognize inferential probe trials as novel pairings of stimuli (Shohamy and Wagner, 2008). This finding suggests that details about directly experienced events may sometimes be lost in favor of an abstracted, generalized framework that codes consistencies among distinct stimulus-response relationships. However, whether a similar loss of detailed event information is typical in other inference paradigms, especially those that utilize rapid acquisition procedures (e.g., single-trial learning), is not known. More research is needed to understand how the processes that support inference are related to those implicated in the formation and use of schemas. Consideration of how task dynamics influence the type of representational structure formed may provide important insights into how the hippocampus codes overlapping event information and interacts with other brain regions (in particular, prefrontal cortex) to support mnemonic flexibility.

### SLEEP-BASED REPLAY AS A MECHANISM FOR MEMORY INTEGRATION

Recent theories suggest that hippocampally mediated replay of event sequences during sleep (Hoffman and McNaughton, 2002; Ji and Wilson, 2007) provides a potential mechanism for constructing networks of related memories that anticipate future decisions and actions (Diekelmann and Born, 2010; Sara, 2010; Lewis and Durrant, 2011)—a process referred to as prospective consolidation (Buckner, 2010). Such theories propose that by reactivating memories during sleep, representations are recombined and recoded, resulting in rich networks of related memories that extend beyond initially encoded events. According to this view, stored memories are not veridical representations of events, but rather derived representations formed in anticipation of future use. Sleep-based replay of hippocampal memory traces, therefore, could enhance performance on inference tasks that tap knowledge about the relationship between overlapping events experienced at different times.

The majority of inferential reasoning studies reviewed here administered the training and critical test phases within the same experimental session, leaving open important questions about the impact of sleep on inference performance. However, one study using the transitive inference paradigm directly examined whether or not sleep enhances inferential reasoning ability (Ellenbogen et al., 2007). Participants were trained on five overlapping associations ( $A > B$ ,  $B > C$ ,  $C > D$ ,  $D > E$ ,  $E > F$ ) and were then tested on novel inferential pairs (e.g.,  $B ? D$ ,  $B ? E$ ) and novel non-inferential pairs ( $A ? F$ ) after a delay of 20 min, 12 h (with or without sleep), or 24 h. While performance on the novel inferential judgments was at chance after a 20-min delay, significant inference was observed at the 12-h (both sleep and wake groups) and 24-h time intervals. Interestingly, the group who performed the inference test after a 12-h period with sleep showed superior performance on the inferential judgments involving two degrees of separation ( $B ? E$ ) compared

to the group who performed the test after a 12-h period of wakefulness.

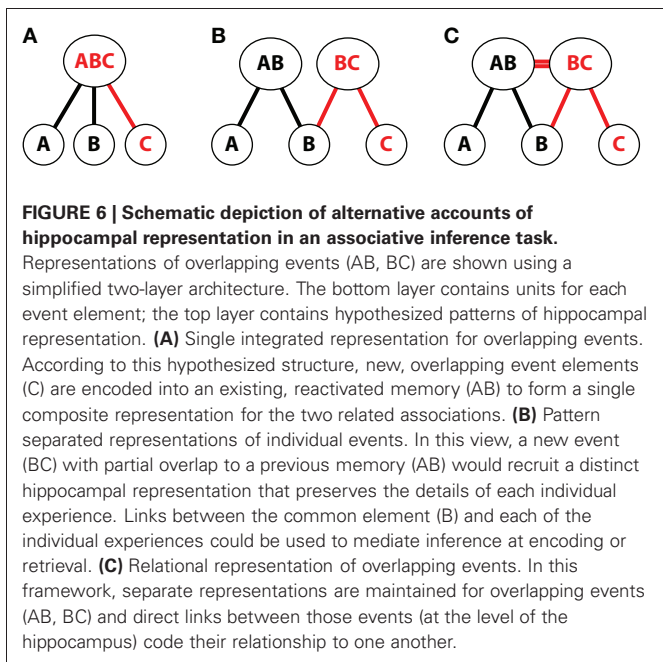
While these findings provide speculative evidence that sleep facilitates inference, it is important to note that the training procedures in this study led to a pattern of behavioral performance that is atypical among animal and human research using this paradigm. Specifically, the study was designed to ensure low levels of performance immediately following learning; thus, relatively weak representations for explicitly learned associations may account for many of the observed effects of sleep. In addition, the time-dependent improvement after the 12-h interval with sleep was not unique to transitive inference judgments but was also observed for the novel non-inferential judgments containing the end elements (A ? F). Therefore, the question of whether sleep enhances memory integration specifically or whether it contributes to improved inferential judgments by simply consolidating memories for individual premise associations is yet to be determined. Future research is needed to fully establish the precise role of sleep and the putative neural mechanisms of sleep-based memory integration in the broad range of inferential reasoning tasks described here.

### HIPPOCAMPAL REPRESENTATIONS UNDERLYING INTEGRATION OF MULTIPLE EXPERIENCES

Initial research suggests that one way in which the hippocampus contributes to inferential reasoning is by integrating information across multiple experiences to establish links between related events, either when new experiences are first learned or offline through replay of related experiences during sleep. However, intriguing questions remain regarding the precise nature of the underlying hippocampal representations.

Several theoretical and computational frameworks have proposed alternate accounts of the properties of memory representations that can support inference. One hypothesized representational structure supporting inference across experiences is one in which new events are incorporated into existing memory traces to be parsimoniously represented in a single, composite memory representation (**Figure 6A**). For instance, consider the simplified example of two events that share a common element (AB, BC) as used in the associative inference paradigm. When a new event occurs that contains an element overlapping with a previous event (e.g., BC after encoding of AB), the overlapping element (B) can trigger pattern completion of the previously encoded memory (AB). According to this hypothesized representational structure, elements from the new, overlapping event (in this case, C) would be encoded *into* the existing, reactivated memory (AB) to form a single integrated representation that combines the two experiences (ABC). Because these integrated representations directly code the novel relationship between A and C along with the original experiences, this representational format provides a basis for the inferential use of memory, but has a notable cost in that details of the individual experiences may not be preserved (e.g., the knowledge that A and C were presented in two different temporal contexts).

The influential cognitive map theory (Tolman, 1948; O'Keefe and Nadel, 1978)—which first sparked interest in the inferential function of the hippocampus—implicitly assumes such



integrated representations. In the context of this theoretical spatial framework, memory traces for newly learned individual events (i.e., recently traveled routes) are combined with memories of previously traveled routes to allow for the creation of an integrated map of the environment, including information about paths not traveled. As a cognitive map of an environment becomes established, it can be reactivated when an animal enters the same environment at a later point and updated with new experiences in that environment. When familiar routes to a goal are blocked, the cognitive map will enable navigation to the goal via an alternate route because information about this novel (i.e., never before traveled) route is included in a single representational structure of the environment. In the context of non-spatial inference tasks, there is some evidence to support this hypothesized representational structure. For example, one study showed that successful participants perform as quickly on inferential judgments as on explicitly trained associations (Shohamy and Wagner, 2008), suggesting similar representations for both directly learned and inferential associations. Moreover, informal assessment suggested that the majority of participants in this study failed to recognize the inferential probes as novel combinations of items, perhaps indicating that some contextual details of original experiences were lost. Returning to our Great Dane example, you may remember that the man and the woman are a couple with a dog, but may not remember specific details about how you first encountered them. Future studies may provide a more detailed account of the circumstances under which memory for original experience may become degraded.

The loss of experiential detail is a significant downside to the single, composite representational structure linking elements of discrete events. Other computational perspectives propose a different representational structure for hippocampus, with pattern separation processes preserving distinct individual experiences and recurrent connections between the element and event

representations allowing inference across experiences (**Figure 6B**; McClelland et al., 1995; Kumaran and McClelland, 2010). In our simplified example, this representational structure would predict that a new event partially overlapping with a previous event (i.e., BC) would recruit a different hippocampal representation to make it distinct from the originally experienced event (AB). The two events would be linked through their individual connections to the shared event element (B). Because of the recurrent connections between individual element and event representations (ascribed to entorhinal cortex and hippocampus, respectively), such a hypothesized structure allows for preservation of event details while also supporting inferential judgments about the relationship between experiences. For example, when presented with a novel inferential probe (AC), each individual element (A and C) may serve as a partial cue leading to the reactivation of the originally experienced events (AB and BC). Activation of the common item (B) in both cases would lead to successful inference. Results showing unique hippocampal responses during inferential retrieval (Preston et al., 2004; Zalesak and Heckers, 2009; Zeithamova and Preston, 2010) might reflect the use of such pattern-separated inputs to support performance. This representational structure can also explain recruitment of the hippocampus during encoding of overlapping events (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010), which potentially reflects changes in the weights linking common elements to the individually experienced events.

It is important to note that even such pattern-separated representations would be expected to change over time and become more generalized. Reactivation of these memory representations during the consolidation process or during sleep-based replay would result in more frequent reactivation of common elements and strengthening of their connections to event representations. In contrast, idiosyncratic elements unique to individual events would be reactivated less frequently and gradually lose their connections to event representations (Lewis and Durrant, 2011). This process would lead to the gradual loss of episodic details in favor of abstracted representations that capture regularities across experiences.

An alternative view that combines elements of both of these frameworks stems from relational memory theory (Cohen and Eichenbaum, 1993). Relational memory theory proposes that the hippocampus maintains representations of individual events while also directly encoding relationships between separate experiences (Eichenbaum, 1999). In our symbolic representation of this theory, different hippocampal units are recruited to represent individual events, but a lateral connection exists at the second level, linking the representations of overlapping events together (**Figure 6C**). Both pattern separation and pattern completion at the level of the hippocampus would contribute to the formation of such networks of related memories. For example, a new overlapping event (BC) would recruit a hippocampal representation distinct from the originally experienced event (AB). Simultaneously, the overlapping element (B) serves as a partial cue that reactivates the prior event (AB). Based on a Hebbian learning rule, the connection between the two hippocampal memory traces would be strengthened and an explicit link between the overlapping events would be formed. Like the

representational structure above, such relational networks would support mnemonic inference while simultaneously preserving memory for individual experiences. It remains for empirical research to test the predictions stemming from these alternate theories of hippocampal representation. In doing so, we will gain a more complete understanding of the computational properties of the hippocampus underlying inferential reasoning, and whether or not the same representational format is used across all inference paradigms, training procedures, and individuals (see Kumaran, 2012 for further discussion).

## CONCLUDING REMARKS

In summary, extensive evidence indicates that the hippocampus plays an important role in inferential reasoning abilities that require judgments about the relationship between multiple items or events. It does so by building flexible memory representations that provide details not only about individual event elements, but also about the relationships between different events. In this way, the function of the hippocampus is not merely to enable the retrospective use of memory; rather, hippocampal function is “intrinsically prospective” (Klein et al., 2002), aimed at constructing representations that can be used to successfully negotiate future judgments and actions. Inferential reasoning tasks thus provide a powerful tool for studying this adaptive nature of memory and how the computational properties of the hippocampus allow memories to be reconstructed into prospectively useful formats.

More generally, the findings reviewed here add to a growing body of evidence that hippocampal processing and representation play an important role in behaviors beyond the episodic memory domain, including working memory (Ranganath and D'Esposito, 2001; Hannula et al., 2006; Olson et al., 2006a,b), implicit memory (Chun and Phelps, 1999; Ryan et al., 2000; Daselaar et al., 2006; Schnyer et al., 2006; Greene et al., 2007; Preston and Gabrieli, 2008; Hannula and Greene, 2012), perception (Barense et al., 2005; Lee et al., 2005a,b; Barense et al., 2007; Bussey and Saksida, 2007; Lee et al., 2012), and even short-term planning and imagining of future events (Addis et al., 2007; Hassabis et al., 2007a,b; Addis and Schacter, 2008, 2011; Voss et al., 2011; Johnson et al., 2012). Such findings do not fit within standard memory systems theories that focus on task-based dissociations stemming from the verbalizability of information (Squire, 1992; Gabrieli, 1998) and require new conceptions of hippocampal function. The research outlined here emphasizes the need to center research efforts on specific hippocampal computations that may serve a number of behaviors, thus providing a theoretical and empirical basis for the broader role of the hippocampus in cognition.

## ACKNOWLEDGMENTS

Supported by a National Science Foundation CAREER Award (Alison R. Preston), Army Research Office Grant 55830-LS-YIP (Alison R. Preston), the National Alliance for Research on Schizophrenia and Depression (Alison R. Preston), NIH-NIMH National Research Service Award F32MH094085 (Dagmar Zeithamova), and the National Defense Science and Engineering Graduate Fellowship Program (Margaret L. Schlichting).



## REFERENCES

- Acuna, B. D., Eliassen, J. C., Donoghue, J. P., and Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cereb. Cortex* 12, 1312–1321.
- Addis, D. R., and Schacter, D. L. (2008). Constructive episodic simulation: temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus* 18, 227–237.
- Addis, D. R., and Schacter, D. L. (2011). The hippocampus and imagining the future: where do we stand? *Front. Hum. Neurosci.* 5:173. doi: 10.3389/fnhum.2011.00173
- Addis, D. R., Wong, A. T., and Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45, 1363–1377.
- Bakker, A., Kirwan, C. B., Miller, M., and Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* 319, 1640–1642.
- Barense, M. D., Bussey, T. J., Lee, A. C., Rogers, T. T., Davies, R. R., Saksida, L. M., Murray, E. A., and Graham, K. S. (2005). Functional specialization in the human medial temporal lobe. *J. Neurosci.* 25, 10239–10246.
- Barense, M. D., Gaffan, D., and Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia* 45, 2963–2974.
- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge, England: Cambridge University Press.
- Buckmaster, C. A., Eichenbaum, H., Amaral, D. G., Suzuki, W. A., and Rapp, P. R. (2004). Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. *J. Neurosci.* 24, 9811–9825.
- Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annu. Rev. Psychol.* 61, 27–48, C21–C28.
- Bunsey, M., and Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature* 379, 255–257.
- Bussey, T. J., and Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: thinking outside of the boxes. *Hippocampus* 17, 898–908.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., and Gabrieli, J. D. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149.
- Chun, M. M., and Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat. Neurosci.* 2, 844–847.
- Cohen, N. J., and Eichenbaum, H. E. (1993). *Memory, Amnesia, and the Hippocampal System*. Cambridge, MA: The MIT Press.
- Coutureau, E., Killcross, A. S., Good, M., Marshall, V. J., Ward-Robinson, J., and Honey, R. C. (2002). Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. *J. Exp. Psychol. Anim. Behav. Process.* 28, 388–396.
- Daselaar, S. M., Fleck, M. S., Prince, S. E., and Cabeza, R. (2006). The medial temporal lobe distinguishes old from new independently of consciousness. *J. Neurosci.* 26, 5835–5839.
- DeVito, L. M., Kanter, B. R., and Eichenbaum, H. (2010a). The hippocampus contributes to memory expression during transitive inference in mice. *Hippocampus* 20, 208–217.
- DeVito, L. M., Lykken, C., Kanter, B. R., and Eichenbaum, H. (2010b). Prefrontal cortex: role in acquisition of overlapping associations and transitive inference. *Learn. Mem.* 17, 161–167.
- Diekelmann, S., and Born, J. (2010). The memory function of sleep. *Nat. Rev. Neurosci.* 11, 114–126.
- Dusek, J. A., and Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7109–7114.
- Eichenbaum, H. (1999). The hippocampus and mechanisms of declarative memory. *Behav. Brain Res.* 103, 123–133.
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* 1, 41–50.
- Eichenbaum, H., and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. New York, NY: Oxford University Press.
- Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., and Tanila, H. (1999). The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron* 23, 209–226.
- Eichenbaum, H., Schoenbaum, G., Young, B., and Bunsey, M. (1996). Functional organization of the hippocampal memory system. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13500–13507.
- Eichenbaum, H., Yonelinas, A. P., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.* 30, 123–152.
- Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., and Walker, M. P. (2007). Human relational memory requires time and sleep. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7723–7728.
- Frank, M. J., Rudy, J. W., Levy, W. B., and O'Reilly, R. C. (2005). When logic fails: implicit transitive inference in humans. *Mem. Cognit.* 33, 742–750.
- Frank, M. J., Rudy, J. W., and O'Reilly, R. C. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. *Hippocampus* 13, 341–354.
- Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annu. Rev. Psychol.* 49, 87–115.
- Greene, A. J., Gross, W. L., Elsinger, C. L., and Rao, S. M. (2006). An fMRI analysis of the human hippocampus: inference, context, and task awareness. *J. Cogn. Neurosci.* 18, 1156–1173.
- Greene, A. J., Gross, W. L., Elsinger, C. L., and Rao, S. M. (2007). Hippocampal differentiation without recognition: an fMRI analysis of the contextual cueing task. *Learn. Mem.* 14, 548–553.
- Gupta, A. S., van der Meer, M. A., Touretzky, D. S., and Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron* 65, 695–705.
- Hannula, D. E., and Greene, A. J. (2012). The hippocampus reevaluated in automatic and unconscious expressions of memory: at a tipping point? *Front. Hum. Neurosci.* (in press).
- Hannula, D. E., and Ranganath, C. (2008). Medial temporal lobe activity predicts successful relational memory binding. *J. Neurosci.* 28, 116–124.
- Hannula, D. E., Tranel, D., and Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *J. Neurosci.* 26, 8352–8359.
- Hassabis, D., Kumaran, D., and Maguire, E. A. (2007a). Using imagination to understand the neural basis of episodic memory. *J. Neurosci.* 27, 14365–14374.
- Hassabis, D., Kumaran, D., Vann, S. D., and Maguire, E. A. (2007b). Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1726–1731.
- Heckers, S., Zalesak, M., Weiss, A. P., Dittman, T., and Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus* 14, 153–162.
- Hoffman, K. L., and McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science* 297, 2070–2073.
- Iordanova, M. D., Killcross, A. S., and Honey, R. C. (2007). Role of the medial prefrontal cortex in acquired distinctiveness and equivalence of cues. *Behav. Neurosci.* 121, 1431–1436.
- Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107.
- Johnson, A., Varberg, Z., Benhardus, J., Maahs, A., and Schrater, P. (2012). The hippocampus and exploration: dynamically evolving behavior and neural representations. *Front. Hum. Neurosci.* (in press).
- Karlsson, M. P., and Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. *Nat. Neurosci.* 12, 913–918.
- Klein, S. B., Cosmides, L., Tooby, J., and Chance, S. (2002). Decisions and the evolution of memory: multiple systems, multiple functions. *Psychol. Rev.* 109, 306–329.
- Kuhl, B. A., Shah, A. T., Dubrow, S., and Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* 13, 501–506.
- Kumaran, D. (2012). The hippocampus, generalization, and structure learning: an empirical and theoretical perspective. *Front. Hum. Neurosci.* (in press).
- Kumaran, D., and McClelland, J. L. (2010). Generalization through recurrence: an interactive model of the hippocampus. *Soc. Neurosci.* [Online].
- Kumaran, D., Summerfield, J. J., Hassabis, D., and Maguire, E. A. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron* 63, 889–901.
- Lee, A. C., Buckley, M. J., Pegman, S. J., Spiers, H., Scallion, V. L., Gaffan, D., Bussey, T. J., Davies, R. R., Kapur, N., Hodges, J. R., and Graham, K. S. (2005a). Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus* 15, 782–797.
- Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur,

- N., Hodges, J. R., and Graham, K. S. (2005b). Perceptual deficits in amnesia: challenging the medial temporal lobe “mnemonic” view. *Neuropsychologia* 43, 1–11.
- Lee, A. C. H., Yeung, L.-K., and Barense, M. D. (2012). The hippocampus and visual perception. *Front. Hum. Neurosci.* (in press).
- Lewis, P. A., and Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends Cogn. Sci.* 15, 343–351.
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 262, 23–81.
- McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457.
- Moses, S. N., Villate, C., and Ryan, J. D. (2006). An investigation of learning strategy supporting transitive inference performance in humans compared to other species. *Neuropsychologia* 44, 1370–1387.
- Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., Golomb, J., Schnirman, G., and Schwartz, R. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J. Cogn. Neurosci.* 15, 185–193.
- Nagode, J. C., and Pardo, J. V. (2002). Human hippocampal activation during transitive inference. *Neuroreport* 13, 939–944.
- O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. London: Clarendon.
- O'Reilly, R. C., and Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol. Rev.* 108, 311–345.
- Olson, I. R., Moore, K. S., Stark, M., and Chatterjee, A. (2006a). Visual working memory is impaired when the medial temporal lobe is damaged. *J. Cogn. Neurosci.* 18, 1087–1097.
- Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., and Verfaellie, M. (2006b). Working memory for conjunctions relies on the medial temporal lobe. *J. Neurosci.* 26, 4596–4601.
- Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. London: Oxford University Press.
- Preston, A. R., Bornstein, A. M., Hutchinson, J. B., Gaare, M. E., Glover, G. H., and Wagner, A. D. (2010). High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. *J. Cogn. Neurosci.* 22, 156–173.
- Preston, A. R., and Gabrieli, J. D. (2008). Dissociation between explicit memory and configural memory in the human medial temporal lobe. *Cereb. Cortex* 18, 2192–2207.
- Preston, A. R., Shrager, Y., Dudukovic, N. M., and Gabrieli, J. D. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* 14, 148–152.
- Preston, A. R., and Wagner, A. D. (2007). “The medial temporal lobe and memory,” in *Neurobiology of Learning and Memory*, 2nd Edn, eds R. P. Kesner and J. L. Martinez (Oxford: Elsevier), 305–337.
- Ranganath, C. (2010). Binding items and contexts: the cognitive neuroscience of episodic memory. *Curr. Dir. Psychol. Sci.* 19, 131–137.
- Ranganath, C., and D'Esposito, M. (2001). Medial temporal lobe activity associated with active maintenance of novel information. *Neuron* 31, 865–873.
- Robin, N., and Holyoak, K. J. (1995). “Relational complexity and the functions of the prefrontal cortex,” in *The Cognitive Neurosciences*, ed. M. S. Gazzaniga. (Cambridge, MA: MIT Press), 987–997.
- Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychol. Sci.* 11, 454–461.
- Sara, S. J. (2010). Reactivation, retrieval, replay and reconsolidation in and out of sleep: connecting the dots. *Front. Behav. Neurosci.* 4:185. doi: 10.3389/fnbeh.2010.00185
- Schnyer, D. M., Dobbins, I. G., Nicholls, L., Schacter, D. L., and Verfaellie, M. (2006). Rapid response learning in amnesia: delineating associative learning components in repetition priming. *Neuropsychologia* 44, 140–149.
- Shohamy, D., and Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* 60, 378–389.
- Singer, A. C., Karlsson, M. P., Nathe, A. R., Carr, M. F., and Frank, L. M. (2010). Experience-dependent development of coordinated hippocampal spatial activity representing the similarity of related locations. *J. Neurosci.* 30, 11586–11604.
- Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* 99, 195–231.
- Squire, L. R., Stark, C. E., and Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208.
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., and Morris, R. G. (2007). Schemas and memory consolidation. *Science* 316, 76–82.
- Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., and Morris, R. G. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science* 333, 891–895.
- van der Jeugd, A., Goddyn, H., Laeremans, A., Arckens, L., D'Hooge, R., and Verguts, T. (2009). Hippocampal involvement in the acquisition of relational associations, but not in the expression of a transitive inference task in mice. *Behav. Neurosci.* 123, 109–114.
- von Fersen, L., Wynne, C. D. L., Delius, J. D., and Staddon, J. E. R. (1991). Transitive inference in pigeons. *J. Exp. Psychol. Anim. Behav. Process.* 17, 334–341.
- Voss, J. L., Warren, D. E., Gonsalves, B. D., Federmeier, K. D., Tranel, D., and Cohen, N. J. (2011). Spontaneous revisitation during visual exploration as a link among strategic behavior, learning, and the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 108, E402–E409.
- Wallenstein, G. V., Eichenbaum, H., and Hasselmo, M. E. (1998). The hippocampus as an associator of discontiguous events. *Trends Neurosci.* 21, 317–323.
- Wendelken, C., and Bunge, S. A. (2010). Transitive inference: distinct contributions of rostrolateral prefrontal cortex and the hippocampus. *J. Cogn. Neurosci.* 22, 837–847.
- Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature* 397, 613–616.
- Zalesak, M., and Heckers, S. (2009). The role of the hippocampus in transitive inference. *Psychiatry Res.* 172, 24–30.
- Zeineh, M. M., Engel, S. A., Thompson, P. M., and Bookheimer, S. Y. (2003). Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science* 299, 577–580.
- Zeithamova, D., and Preston, A. R. (2010). Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *J. Neurosci.* 30, 14676–14684.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 October 2011; accepted: 13 March 2012; published online: 26 March 2012.

Citation: Zeithamova D, Schlichting ML and Preston AR (2012) The hippocampus and inferential reasoning: building memories to navigate future decisions. *Front. Hum. Neurosci.* 6:70. doi: 10.3389/fnhum.2012.00070

Copyright © 2012 Zeithamova, Schlichting and Preston. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.